

(19)



(11)

EP 1 354 276 B1

(12)

EUROPEAN PATENT SPECIFICATION

(45) Date of publication and mention
of the grant of the patent:
12.12.2007 Bulletin 2007/50

(51) Int Cl:
G10L 19/00 (2006.01)

(21) Application number: **01985137.7**

(86) International application number:
PCT/US2001/050295

(22) Date of filing: **26.10.2001**

(87) International publication number:
WO 2002/037316 (10.05.2002 Gazette 2002/19)

(54) METHOD AND APPARATUS FOR CREATING A UNIQUE AUDIO SIGNATURE

VERFAHREN UND VORRICHTUNG ZUM ERZEUGEN VON EINDEUTIGEN AUDIOSIGNATUREN
PROCEDE ET APPAREIL DE CREATION D'UNE SIGNATURE AUDIO UNIQUE

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU
MC NL PT SE TR

- **BLUM, Thomas, L.**
San Francisco, CA 94109 (US)
- **KEISLAR, Douglas, F.**
Berkeley, CA 94708 (US)
- **WHEATON, James, A.**
Fairfax, CA 94930 (US)

(30) Priority: **03.11.2000 US 706227**

(43) Date of publication of application:
22.10.2003 Bulletin 2003/43

(73) Proprietor: **Audible Magic Corporation**
Los Gatos, CA 95032 (US)

(74) Representative: **Leeming, John Gerard**
J.A. Kemp & Co.
14 South Square
Gray's Inn
London WC1R 5JJ (GB)

(72) Inventors:
• **WOLD, Erling, H.**
El Cerrito, CA 94530 (US)

(56) References cited:
US-A- 5 210 820 **US-A- 5 983 176**
US-A- 6 096 961

EP 1 354 276 B1

Note: Within nine months from the publication of the mention of the grant of the European patent, any person may give notice to the European Patent Office of opposition to the European patent granted. Notice of opposition shall be filed in a written reasoned statement. It shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

Description

[0001] The present invention relates to data communications. In particular, the present invention relates to creating a unique audio signature.

[0002] Digital audio technology has greatly changed the landscape of music and entertainment. Rapid increases in computing power coupled with decreases in cost have made it possible individuals to generate finished products having a quality once available only in a major studio. Once consequence of modern technology is that legacy media storage standards, such as reel-to-reel tapes, are being rapidly replaced by digital storage media, such as the Digital Versatile Disk (DVD), and Digital Audio Tape (DAT). Additionally, with higher capacity hard drives standard on most personal computers, home users may now store digital files such as audio or video tracks on their home computers.

[0003] Furthermore, the Internet has generated much excitement, particularly among those who see the Internet as an opportunity to develop new avenues for artistic expression and communication. The Internet has become a virtual gallery, where artists may post their works on a Web page. Once posted, the works may be viewed by anyone having access to the Internet.

[0004] One application of the Internet that has received considerable attention is the ability to transmit recorded music over the Internet. Once music has been digitally encoded into a file, the file may be both downloaded by users for play, or broadcast ("streamed") over the Internet. When files are streamed, they may be listened to by Internet users in a manner much like traditional radio stations.

[0005] Given the widespread use of digital media, digital audio files, or digital video files containing audio information, may need to be identified. The need for identification of digital files may arise in a variety of situations. For example, an artist may wish to verify royalty payments or generate their own Arbitron®-like ratings by identifying how often their works are being streamed or downloaded. Additionally, users may wish to identify a particular work. The prior art has made efforts to create methods for identifying digital audio works.

[0006] However, systems of the prior art suffer from certain disadvantages. For example, prior art systems typically create a reference signature by examining the copyrighted work as a whole, and then creating a signature based upon the audio characteristics of the entire work. However, examining a work in total can result in a signature may not accurately represent the original work. Often, a work may have distinctive passages which may not be reflected in a signature based upon the total work. Furthermore, often works are electronically processed prior to being streamed or downloaded, in a manner that may affect details of the work's audio characteristics, which may result in prior art systems missing the identification of such works. Examples of such electronic processing include data compression and various sorts of audio signal processing such as equalization.

[0007] US 5,918,223 discloses a method for determining a work in which the work is segmented, a signature generated and compared to a reference signature to determine if the work is known.

[0008] Hence, there exists a need to provide a system which overcomes the disadvantages of the prior art.

[0009] The present invention relates to data communications. In particular, the present invention relates to creating a unique audio signature.

[0010] According to a first aspect of the invention there is provided a method for determining an identity of a sampled work, said method comprising receiving data of a sampled work, segmenting said data of said sampled work into a plurality of segments wherein each of said segments has predetermined segment size and a predetermined hop size, creating a signature of said sample work based upon said plurality of segments, comparing said signature of said sampled work to a plurality of signatures of reference works, and determining said sampled work is one of said reference works based upon said comparison, said method characterized in that said predetermined hop size of said segments of said sampled work signature is chosen to be less than said hop size of each of said plurality of reference signatures.

[0011] According to a further aspect of the invention there is provided an apparatus that determines an identity of a sampled work, said apparatus comprising circuitry configured to receive data of a sampled work, circuitry configured to segment said data of said sampled work into a plurality of segments wherein each of said segments has predetermined segment size and a predetermined hop size, circuitry configured to create a signature of said sampled work based upon said plurality of segments, circuitry configured to compare said signature of said sampled work to a plurality of signatures of reference works, and circuitry configured to determine said sampled work is one of said reference works based upon said comparison, said apparatus characterized in that said predetermined hop size of said segments of said sampled work signature is chosen to be less than said hop size of each of said plurality of reference signatures.

[0012] According to a further aspect of the invention there is provided a program storage device readable by a machine, tangibly embodying a program of instructions executable by the machine to perform a method as described above.

Figure 1 is a flowchart of a method according to the present invention.

Figure 2 is a diagram of a system suitable for use with the present invention.

Figure 3 is a diagram of segmenting according to the present invention.

Figure 4 is a detailed diagram of segmenting according to the present invention showing hop size.

Figure 5 is a graphical flowchart showing the creating of a segment feature vector according to the present invention.

Figure 6 is a diagram of a signature according to the present invention.

Figure 7 is a functional diagram of a comparison process according to the present invention.

[0013] Persons of ordinary skill in the art will realize that the following description of the present invention is illustrative only and not in any way limiting. Other embodiments of the invention will readily suggest themselves to such skilled persons having the benefit of this disclosure.

[0014] It is contemplated that the present invention may be embodied in various computer and machine-readable data structures. Furthermore, it is contemplated that data structures embodying the present invention will be transmitted across computer and machine-readable media, and through communications systems by use of standard protocols such as those used to enable the Internet and other computer networking standards.

[0015] The invention further relates to machine-readable media on which are stored embodiments of the present invention. It is contemplated that any media suitable for storing instructions related to the present invention is within the scope of the present invention. By way of example, such media may take the form of magnetic, optical, or semiconductor media.

[0016] The present invention may be described through the use of flowcharts. Often, a single instance of an embodiment of the present invention will be shown. As is appreciated by those of ordinary skill in the art, however, the protocols, processes, and procedures described herein may be repeated continuously or as often as necessary to satisfy the needs described herein. Accordingly, the representation of the present invention through the use of flowcharts should not be used to limit the scope of the present invention.

[0017] The present invention may also be described through the use of web pages in which embodiments of the present invention may be viewed and manipulated. It is contemplated that such web pages may be programmed with web page creation programs using languages standard in the art such as HTML or XML. It is also contemplated that the web pages described herein may be viewed and manipulated with web browsers running on operating systems standard in the art, such as the Microsoft Windows® and Macintosh® versions of Internet Explorer® and Netscape®. Furthermore, it is contemplated that the functions performed by the various web pages described herein may be implemented through the use of standard programming languages such as Java® or similar languages.

[0018] The present invention will first be described in general overview. Then, each element will be described in further detail below.

[0019] Referring now to Figure 1, a flowchart is shown which provides a general overview of the present invention. The present invention may be viewed as three steps: 1) receiving a sampled work; 2) segmenting the work; 3) creating signatures of the segments; and 4) storing the signatures of the segments.

Receiving a sampled work

[0020] Beginning with act 100, a sampled work is provided to the present invention. It is contemplated that the work will be provided to the present invention as a digital audio stream.

[0021] It should be understood that if the audio is in analog form, it may be digitized in a manner standard in the art.

Segmenting the work

[0022] After the sampled work is received, the work is then segmented in act 102. It is contemplated that the sampled work may be segmented into predetermined lengths. Though segments may be of any length, the segments of the present invention are preferably of the same length.

[0023] In an exemplary non-limiting embodiment of the present invention, the segment lengths are in the range of 0.5 to 3 seconds. It is contemplated that if one were searching for very short sounds (e.g., sound effects such as gunshots), segments as small as 0.01 seconds may be used in the present invention. Since humans don't resolve audio changes below about 0.018 seconds, segment lengths less than 0.018 seconds may not be useful. On the other hand, segment lengths as high as 30-60 seconds may be used in the present invention. The inventors have found that beyond 30-80 seconds may not be useful, since most details in the signal tend to average out.

Generating signatures

[0024] Next, in act 104, each segment is analyzed to produce a signature, known herein as a segment feature vector.

It is contemplated that a wide variety of methods known in the art may be used to analyze the segments and generate segment feature vectors. In an exemplary non-limiting embodiment of the present invention, the segment feature vectors may be created using the method described in US Patent #5,918,223 to Blum.

Storing the signatures

[0025] In act 106, the segment feature vectors are stored to create a representative signature of the sampled work.

[0026] Each above-listed step will now be shown and described in detail.

5 [0027] Referring now to Figure 2, a diagram of a system suitable for use with the present invention is shown. FIG. 2 includes a client system 200. It is contemplated that client system 200 may comprise a personal computer 202 including hardware and software standard in the art to run an operating system such as Microsoft Windows®, MAC OS®, or other operating systems standard in the art. Client system 200 may further include a database 204 for storing and retrieving
10 embodiments of the present invention. It is contemplated that database 204 may comprise hardware and software standard in the art and may be operatively coupled to PC 202. Database 204 may also be used to store and retrieve the works and segments utilized by the present invention.

[0028] Client system 200 may further include an audio/video (A/V) input device 208. A/V device 208 is operatively coupled to PC 202 and is configured to provide works to the present invention which may be stored in traditional audio or video formats. It is contemplated that A/V device 208 may comprise hardware and software standard in the art
15 configured to receive and sample audio works (including video containing audio information), and provide the sampled works to the present invention as digital audio files. Typically, the A/V input device 208 would supply raw audio samples in a format such as 16-bit stereo PCM format. A/V input device 208 provides an example of means for receiving a sampled work.

[0029] It is contemplated that sampled works may be obtained over the Internet, also. Typically, streaming media over the Internet is provided by a provider, such as provider 218 of FIG. 2. Provider 218 includes a streaming application server 220, configured to retrieve works from database 222 and stream the works in a format standard in the art, such as Real®, Windows Media®, or QuickTime®. The server then provides the streamed works to a web server 224, which then provides the streamed work to the Internet 214 through a gateway 216. Internet 214 may be any packet-based
20 network standard in the art, such as IP, Frame Relay, or ATM.

[0030] To reach the provider 218, the present invention may utilize a cable or DSL head end 212 standard in the art operatively, which is coupled to a cable modem or DSL modem 210 which is in turn coupled to the system's network 206. The network 206 may be any network standard in the art, such as a LAN provided by a PC 202 configured to run software standard in the art.

[0031] It is contemplated that the sampled work received by system 200 may contain audio information from a variety of sources known in the art, including, without limitation, radio, the audio portion of a television broadcast, Internet radio, the audio portion of an Internet video program or channel, streaming audio from a network audio server, audio delivered to personal digital assistants over cellular or wireless communication systems, or cable and satellite broadcasts.
30

[0032] Additionally, it is contemplated that the present invention may be configured to receive and compare segments coming from a variety of sources either stored or in real-time. For example, it is contemplated that the present invention may compare a real-time streaming work coming from streaming server 218 or A/V device 208 with a reference segment stored in database 204.
35

[0033] Figure 3 shows a diagram showing the segmenting of a work according to the present invention. FIG. 3 includes audio information 300 displayed along a time axis 302. FIG. 3 further includes a plurality of segments 304, 306, and 308 taken of audio information 300 over some segment size T.
40

[0034] In an exemplary non-limiting embodiment of the present invention, instantaneous values of a variety of acoustic features are computed at a low level, preferably about 100 times a second. Additionally, 10 MFCCs (cepstral coefficients) are computed for each segment. It is contemplated that any number of MFCCs may be computed. Preferably, 5-20 MFCCs are computed, however, as many as 30 MFCCs may be computed, depending on the need for accuracy versus speed.
45

[0035] In an exemplary non-limiting embodiment of the present invention, the segment-level acoustical features comprise statistical measures as disclosed in the '223 patent of these low-level features calculated over the length of each segment. The data structure may store other bookkeeping information as well (segment size, hop size, item ID, UPC, etc).
50

[0036] As can be seen by inspection of FIG. 3, the segments 304, 306, and 308 may overlap in time. This amount of overlap may be represented by measuring the time between the center point of adjacent segments. This amount of time is referred to herein as the hop size of the segments, and is so designated in FIG. 3. By way of example, if the segment length T of a given segment is one second, and adjacent segments overlap by 50%, the hop size would be 0.5 second.
55

[0037] The hop size may be set during the development of the software. Additionally, the hop sizes of the reference database and the real-time segments may be predetermined to facilitate compatibility. For example, the reference signatures in the reference database may be precomputed with a fixed hop and segment size, and thus the client applications should conform to this segment size and have a hop size which integrally divides the reference signature hop size. It is contemplated that one may experiment with a variety of segment sizes in order to balance the tradeoff of accuracy with speed of computation for a given application.

[0038] The inventors have found that by carefully choosing the hop size of the segments, the accuracy of the identi-

fication process may be significantly increased. Additionally, the inventors have found that the accuracy of the identification process may be increased if the hop size of reference segments and the hop size of segments obtained in real-time are each chosen independently. The importance of the hop size of segments may be illustrated by examining the process for segmenting pre-recorded works and real-time works separately.

Reference signatures

[0039] Prior to attempting to identify a given work, a reference database of signatures must be created. When building a reference database, a segment length having a period of less than three seconds is preferred. In an exemplary non-limiting embodiment of the present invention, the segment lengths have a period ranging from 0.5 seconds to 3 seconds. For a reference database, the inventors have found that a hop size of approximately 50% to 100% of the segment size is preferred.

[0040] It is contemplated that the reference signatures may be stored on a database such as database 204 as described above. Database 204 and the discussion herein provide an example of means for providing a plurality of reference signatures each having a segment size and a hop size.

Real-time signatures

[0041] The choice of the hop size is important for real-time segments.

[0042] Figure 4 shows a detailed diagram of a real-time segment according to the present invention. FIG. 4 includes real-time audio information 400 displayed along a time axis 402. FIG. 4 further includes segments 404 and 406 taken of audio information 400 over some segment length T. In an exemplary non-limiting embodiment of the present invention, the segment length of real-time segments is chosen to range from 0.5 to 3 seconds.

[0043] As can be seen by inspection of FIG. 4, the hop size of real-time is chosen to be smaller than that of reference segments. In an exemplary non-limiting embodiment of the present invention, the hop size of real-time segments is less than 50% of the segment size. In yet another exemplary non-limiting embodiment of the present invention, the real-time hop size may be 0.1 seconds.

[0044] The inventors have found such a small hop size advantageous for the following reasons. The ultimate purpose of generating real-time segments is to analyze and compare them with the reference segments in the database to look for matches. The inventors have found at least two major reasons why a segment of the same audio recording captured real-time would not match its counterpart in the database. One is that the broadcast channel does not produce a perfect copy of the original. For example, the work may be edited or processed or the announcer may talk over part of the work. The other reason is that larger segment boundaries may not line up in time with the original segment boundaries of the target recordings.

[0045] The inventors have found that by choosing a smaller hop size, some of the segments will ultimately have time boundaries that line up with the original segments, notwithstanding the problems listed above. The segments that line up with a "clean" segment of the work may then be used to make an accurate comparison while those that do not so line up may be ignored. The inventors have found that a hop size of 0.1 seconds seems to be the maximum that would solve this time shifting problem.

[0046] As mentioned above, once a work has been segmented, the individual segments are then analyzed to produce a segment feature vector. Figure 5 is a diagram showing an overview of how the segment feature vectors may be created using the methods described in US Patent #5,918,223 to Blum, et al. It is contemplated that a variety of analysis methods may be useful in the present invention, and many different features may be used to make up the feature vector. The inventors have found that the pitch, brightness, bandwidth, and loudness features of the '223 patent to be useful in the present invention. Additionally, spectral features may be used analyzed, such as the energy in various spectral bands. The inventors have found that the cepstral features (MFCCs) are very robust (more invariant) given the distortions typically introduced during broadcast, such as EQ, multi-band compression/limiting, and audio data compression techniques such as MP3 encoding/decoding, etc.

[0047] In act 500, the audio segment is sampled to produce a segment. In act 502, the sampled segment is then analyzed using Fourier Transform techniques to transform the signal into the frequency domain. In act 504, mel frequency filters are applied to the transformed signal to extract the significant audible characteristics of the spectrum. In act 506, a Discrete Cosine Transform is applied which converts the signal into mel frequency cepstral coefficients (MFCCs). Finally, in act 508, the MFCCs are then averaged over a predetermined period. In an exemplary non-limiting embodiment of the present invention, this period is approximately one second. Additionally, other characteristics may be computed at this time, such as brightness or loudness. A segment feature vector is then produced which contains a list containing at least the 10 MFCCs corresponding average.

[0048] The disclosure of FIGS. 3, 4, and 5 provide examples of means for creating a signature of a sampled work having a segment size and a hop size.

[0049] Figure 6 is a diagram showing a complete signature 600 according to the present invention. Signature 600 includes a plurality of segment feature vectors 1 through n generated as shown and described above. Signature 600 may also include an identification portion containing a unique ID. It is contemplated that the identification portion may contain a unique identifier provided by the RIAA (Recording Industry Association of America). The identification portion may also contain information such as the UPC (Universal Product Code) of the various products that contain the audio corresponding to this signature. Additionally, it is contemplated that the signature 600 may also contain information pertaining to the characteristics of the file itself, such as the hop size, segment size, number of segments, etc., which may be useful for storing and indexing.

[0050] Signature 600 may then be stored in a database and used for comparisons. The following computer code in the C programming language provides an example of a database structure in memory according to the present invention:

```
typedef struct
{
    float hopSize;           /* hop size */
    float segmentSize;      /* segment size */
    MFSignature* signatures; /* array of signatures */
} MFDATABASE;
```

[0051] The following provides an example of the structure of a segment according to the present invention:

```
typedef struct
{
    char* id;                /* unique ID for this audio clip */
    long numSegments; /* number of segments */
    float* features;        /* feature array */
    long size;              /* size of per-segment feature vector */
    float hopSize;
    float segmentSize;
} MFSegment;
```

[0052] The discussion of FIG. 6 provides an example of means for storing segments and signatures according to the present invention.

[0053] Figure 7 shows a functional diagram of a comparison process according to the present invention. Act 1 of FIG. 7 shows unknown audio being converted to a signature according to the present invention. In act 2, reference signatures are retrieved from a reference database. Finally, the reference signatures are scanned and compared to the unknown audio signatures to determine whether a match exists. This comparison may be accomplished through means known in the art. For example, the Euclidean distance between the reference and real-time signature can be computed and compared to a threshold.

[0054] It is contemplated that the present invention has many beneficial uses, including many outside of the music piracy area. For example, the present invention may be used to verify royalty payments. The verification may take place at the source or the listener. Also, the present invention may be utilized for the auditing of advertisements, or collecting Arbitron®-like data (who is listening to what). The present invention may also be used to label the audio recordings on a user's hard disk or on the web.

[0055] While embodiments and applications of this invention have been shown and described, it would be apparent to those skilled in the art that many more modifications than mentioned above are possible within the scope defined by the appended claims.

Claims

1. A method for determining an identity of a sampled work, said method comprising receiving data of a sampled work, segmenting said data of said sampled work into a plurality of segments wherein each of said segments has predetermined segment size and a predetermined hop size, creating a signature of said sampled work based upon said plurality of segments, comparing said signature of said sampled work to a plurality of signatures of reference works, and determining said sampled work is one of said reference works based upon said comparison, said method characterized in that said predetermined hop size of said segments of said sampled work signature is chosen to be less than said hop size of each of said plurality of reference signatures.
2. The method of claim 1, wherein said act of creating a signature of said sampled work comprises calculating segment

feature vectors for each segment of said sampled work.

3. The method of claim 1, wherein said act of creating a signature of said sampled work includes calculating a plurality of MFCCs for each said segment.
4. The method of claim 1, wherein said act of creating a signature of said sampled work includes calculating a plurality of acoustical features from the group consisting of at least one of loudness, pitch, brightness, bandwidth, spectrum and MFCC coefficients for each said segment.
5. The method of claim 1, wherein said sampled work signature comprises a plurality of segments and an identification portion.
6. The method of claim 1, wherein said plurality of segments of said sampled work signature comprise a segment size of approximately 0.5 to 3 seconds.
7. The method of claim 6, wherein said plurality of segments of said sampled work signature comprise a hop size of less than 50% of the segment size.
8. The method of claim 6, wherein said plurality of segments of said sampled work signature comprise a hop size of approximately 0.1 seconds.
9. An apparatus that determines an identity of a sampled work, said apparatus comprising circuitry configured to receive data of a sampled work, circuitry configured to segment said data of said sampled work into a plurality of segments wherein each of said segments has predetermined segment size and a predetermined hop size, circuitry configured to create a signature of said sampled work based upon said plurality of segments, circuitry configured to compare said signature of said sampled work to a plurality of signatures of reference works, and circuitry configured to determine said sampled work is one of said reference works based upon said comparison, said apparatus **characterized in that** said predetermined hop size of said segments of said sampled work signature is chosen to be less than said hop size of each of said plurality of reference signatures.
10. The apparatus of claim 9, wherein said circuitry configured to create a signature of said sampled work comprises circuitry configured to calculate segment feature vectors for each of said plurality of segments of said sampled work.
11. The apparatus of claim 9, wherein said circuitry configured to create a signature includes calculating a plurality of MFCCs for each said segment.
12. The apparatus of claim 9, wherein said circuitry configured to create a signature includes circuitry configured to calculate one of a plurality of acoustical features selected from a group consisting of loudness, pitch, brightness, bandwidth, spectrum and MFCC coefficients for each of said plurality of segments of said sampled works.
13. The apparatus of claim 9, wherein said sampled work signature comprises a plurality of segments and an identification portion.
14. The apparatus of claim 9, wherein said plurality of segments of said sampled work comprise said predetermined segment size of approximately 0.5 to 3 seconds.
15. The apparatus of claim 14, wherein said predetermined hop size of said plurality of segments of said sampled work signature is less than 50% of the segment size.
16. The apparatus of claim 14, wherein said predetermined hop size of each of said plurality of segments of said sampled work signature is approximately 0.1 seconds.
17. A program storage device readable by a machine, tangibly embodying a program of instructions executable by the machine to perform a method according to any one of claims 1 to 8.

Patentansprüche

1. Verfahren zum Ermitteln einer Identität eines abgetasteten Werks, wobei das Verfahren umfasst: das Empfangen von Daten eines abgetasteten Werks, das Segmentieren der Daten des abgetasteten Werks in mehrere Segmente, wobei jedes der Segmente eine vorbestimmte Segmentgröße und eine vorbestimmte Hop-Size aufweist, das Erzeugen einer Signatur des abgetasteten Werks beruhend auf den mehreren Segmenten, das Vergleichen der Signatur des abgetasteten Werks mit mehreren Signaturen von Referenzwerken und beruhend auf dem Vergleich das Bestimmen, dass das abgetastete Werk eines der Referenzwerke ist, wobei das Verfahren **dadurch gekennzeichnet ist, dass** die vorbestimmte Hop-Size der Segmente der Signatur des abgetasteten Werks kleiner als die Hop-Size jeder der mehreren Referenzsignaturen gewählt wird.
2. Verfahren nach Anspruch 1, **dadurch gekennzeichnet, dass** der Vorgang des Erzeugens einer Signatur des abgetasteten Werks das Berechnen von Segmentmerkmalsfaktoren für jedes Segment des abgetasteten Werks umfasst.
3. Verfahren nach Anspruch 1, **dadurch gekennzeichnet, dass** der Vorgang des Erzeugens einer Signatur des abgetasteten Werks das Berechnen mehrerer MFCC für jedes Segment umfasst.
4. Verfahren nach Anspruch 1, **dadurch gekennzeichnet, dass** der Vorgang des Erzeugens einer Signatur des abgetasteten Werks das Berechnen mehrerer akustischer Merkmale aus der Gruppe bestehend aus mindestens einem von Lautstärke, Tonhöhe, Helligkeit, Bandbreite, Spektrum und MFCC-Koeffizienten für jedes Segment umfasst.
5. Verfahren nach Anspruch 1, **dadurch gekennzeichnet, dass** die Signatur des abgetasteten Werks mehrere Segmente und einen Erkennungsabschnitt umfasst.
6. Verfahren nach Anspruch 1, **dadurch gekennzeichnet, dass** die mehreren Segmente der Signatur des abgetasteten Werks eine Segmentgröße von in etwa 0,5 bis 3 Sekunden umfassen.
7. Verfahren nach Anspruch 6, **dadurch gekennzeichnet, dass** die mehreren Segmente der Signatur des abgetasteten Werks eine Hop-Size von kleiner als 50% der Segmentgröße umfassen.
8. Verfahren nach Anspruch 6, **dadurch gekennzeichnet, dass** die mehreren Segmente der Signatur des abgetasteten Werks eine Hop-Size von in etwa 0,1 Sekunden umfassen.
9. Vorrichtung, die eine Identität eines abgetasteten Werks ermittelt, wobei die Vorrichtung umfasst: eine zum Empfangen von Daten eines abgetasteten Werks ausgelegte Schaltung, eine zum Segmentieren der Daten des abgetasteten Werks in mehrere Segmente ausgelegte Schaltung, wobei jedes der Segmente eine vorbestimmte Segmentgröße und eine vorbestimmte Hop-Size aufweist, eine zum Erzeugen einer Signatur des abgetasteten Werks beruhend auf den mehreren Segmenten ausgelegte Schaltung, eine zum Vergleichen der Signatur des abgetasteten Werks mit mehreren Signaturen von Referenzwerken ausgelegte Schaltung und eine zum vergleichsbasierten Bestimmen, dass das abgetastete Werk eines der Referenzwerke ist, ausgelegte Schaltung, wobei die Vorrichtung **dadurch gekennzeichnet ist, dass** die vorbestimmte Hop-Size der Segmente der Signatur des abgetasteten Werks kleiner als die Hop-Size jeder der mehreren Referenzsignaturen gewählt ist.
10. Vorrichtung nach Anspruch 9, **dadurch gekennzeichnet, dass** die zum Erzeugen einer Signatur des abgetasteten Werks ausgelegte Schaltung eine zum Berechnen von Segmentmerkmalsvektoren für jedes der mehreren Segmente des abgetasteten Werks ausgelegte Schaltung umfasst.
11. Vorrichtung nach Anspruch 9, **dadurch gekennzeichnet, dass** die zum Erzeugen einer Signatur ausgelegte Schaltung das Berechnen mehrerer MFCC für jedes Segment umfasst.
12. Vorrichtung nach Anspruch 9, **dadurch gekennzeichnet, dass** die zum Erzeugen einer Signatur ausgelegte Schaltung eine Schaltung umfasst, die zum Berechnen eines von mehreren akustischen Merkmalen gewählt aus einer Gruppe bestehend aus Lautstärke, Tonhöhe, Helligkeit, Bandbreite, Spektrum und MFCC-Koeffizienten für jedes der mehreren Segmente der abgetasteten Werke ausgelegt ist.
13. Vorrichtung nach Anspruch 9, **dadurch gekennzeichnet, dass** die Signatur des abgetasteten Werks mehrere

Segmente und einen Erkennungsabschnitt umfasst.

14. Vorrichtung nach Anspruch 9, **dadurch gekennzeichnet, dass** die mehreren Segmente des abgetasteten Werks die vorbestimmte Segmentgröße von in etwa 0,5 bis 3 Sekunden umfassen.
15. Vorrichtung nach Anspruch 14, **dadurch gekennzeichnet, dass** die vorbestimmte Hop-Size der mehreren Segmente der Signatur des abgetasteten Werks kleiner als 50% der Segmentgröße ist.
16. Vorrichtung nach Anspruch 14, **dadurch gekennzeichnet, dass** die vorbestimmte Hop-Size jedes der mehreren Segmente der Signatur des abgetasteten Werks in etwa 0,1 Sekunden ist.
17. Maschinenlesbare Programmspeichervorrichtung, die konkret ein Programm von Befehlen verkörpert, die von der Maschine zum Durchführen eines Verfahrens nach einem der Ansprüche 1 bis 8 ausführbar sind.

Revendications

1. Procédé pour déterminer une identité d'une oeuvre échantillonnée, ledit procédé comportant la réception de données d'une oeuvre échantillonnée, la segmentation desdites données de ladite oeuvre échantillonnée en une pluralité de segments dans lequel chacun desdits segments a une taille de segment prédéterminée et une taille de saut prédéterminée, la création d'une signature de ladite oeuvre échantillonnée sur la base de ladite pluralité de segments, la comparaison de ladite signature de ladite oeuvre échantillonnée à une pluralité de signatures d'oeuvres de référence, et la détermination que ladite oeuvre échantillonnée est l'une desdites oeuvres de référence sur la base de ladite comparaison, ledit procédé étant **caractérisé en ce que** ladite taille de saut prédéterminée desdits segments de ladite signature d'oeuvre échantillonnée est choisie de manière à être inférieure à ladite oeuvre de saut de chacune de ladite pluralité de signatures de référence.
2. Procédé selon la revendication 1, dans lequel ladite action de création de signature de ladite oeuvre échantillonnée comprend le calcul de vecteurs caractéristiques de segment pour chaque segment de ladite oeuvre échantillonnée.
3. Procédé selon la revendication 1, dans lequel ladite action de création de signature de ladite oeuvre échantillonnée inclut le calcul d'une pluralité de MFCC pour chacun desdits segments.
4. Procédé selon la revendication 1, dans lequel ladite action de création de signature de ladite oeuvre échantillonnée inclut le calcul d'une pluralité de caractéristiques acoustiques parmi le groupe constitué d'au moins l'un parmi la sonie, la hauteur tonale, la brillance, la largeur de bande, le spectre et des coefficients MFCC pour chaque segment.
5. Procédé selon la revendication 1, dans lequel ladite signature d'oeuvre échantillonnée comporte une pluralité de segments et une partie d'identification.
6. Procédé selon la revendication 1, dans lequel ladite pluralité de segments de ladite signature d'oeuvre échantillonnée comporte une taille de segment d'approximativement 0,5 à 3 secondes.
7. Procédé selon la revendication 6, dans lequel ladite pluralité de segments de ladite signature d'oeuvre échantillonnée comporte une taille de saut inférieure à 50 % de la taille de segment.
8. Procédé selon la revendication 6, dans lequel ladite pluralité de segments de ladite signature d'oeuvre échantillonnée comporte une taille de saut d'approximativement 0,1 seconde.
9. Appareil qui détermine une identité d'une oeuvre échantillonnée, ledit appareil comportant un circuit configuré pour recevoir des données d'une oeuvre échantillonnée, un circuit configuré pour segmenter lesdites données de ladite oeuvre échantillonnée en une pluralité de segments dans lesquels chacun desdits segments a une taille de segment prédéterminée et une taille de saut prédéterminée, un circuit configuré pour créer une signature de ladite oeuvre échantillonnée sur la base de ladite pluralité de segments, un circuit configuré pour comparer ladite signature de ladite oeuvre échantillonnée à une pluralité de signatures d'oeuvre de référence, et un circuit configuré pour déterminer que ladite oeuvre échantillonnée est l'une desdites oeuvres de référence sur la base de ladite comparaison, ledit appareil étant **caractérisé en ce que** ladite taille de saut prédéterminée desdits segments de ladite signature d'oeuvre échantillonnée est choisie de manière à être inférieure à ladite oeuvre de saut de chacune de ladite pluralité

de signatures de référence.

10. Appareil selon la revendication 9, dans lequel ledit circuit configuré pour créer une signature de ladite oeuvre échantillonnée comprend un circuit configuré pour calculer des vecteurs caractéristiques de segments pour chacun de ladite pluralité de segments de ladite oeuvre échantillonnée.
11. Appareil selon la revendication 9, dans lequel ledit circuit configuré pour créer une signature inclut le calcul d'une pluralité de MFCC pour chaque segment.
12. Appareil selon la revendication 9, dans lequel ledit circuit configuré pour créer une signature inclut un circuit configuré pour calculer l'une d'une pluralité de caractéristiques acoustiques sélectionnées parmi le groupe constitué de la sonie, la hauteur tonale, la brillance, la largeur de bande, le spectre et les coefficients MFCC pour chacun de ladite pluralité de segments desdites oeuvres échantillonnées.
13. Appareil selon la revendication 9, dans lequel ladite signature d'oeuvre échantillonnée comporte une pluralité de segments et une partie d'identification.
14. Appareil selon la revendication 9, dans lequel ladite pluralité de segments de ladite oeuvre échantillonnée comporte ladite taille de segment prédéterminée d'approximativement 0,5 à 3 secondes.
15. Appareil selon la revendication 14, dans lequel ladite taille de saut prédéterminée de ladite pluralité de segments de ladite signature d'oeuvre échantillonnée est inférieure à 50 % de la taille de segment.
16. Appareil selon la revendication 14, dans lequel ladite taille de saut prédéterminée de chacun de ladite pluralité de segments de ladite signature d'oeuvre échantillonnée est approximativement de 0,1 seconde.
17. Dispositif de mémorisation de programme lisible par machine, intégrant de manière tangible un programme d'instructions exécutable par la machine pour exécuter un procédé selon l'une quelconque des revendications 1 à 8.

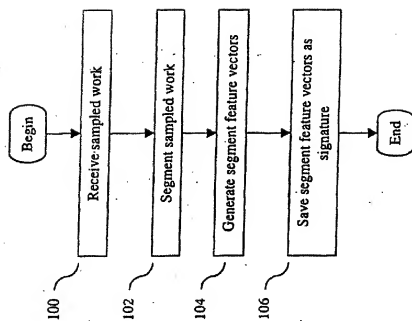


FIG. 1
Present Invention

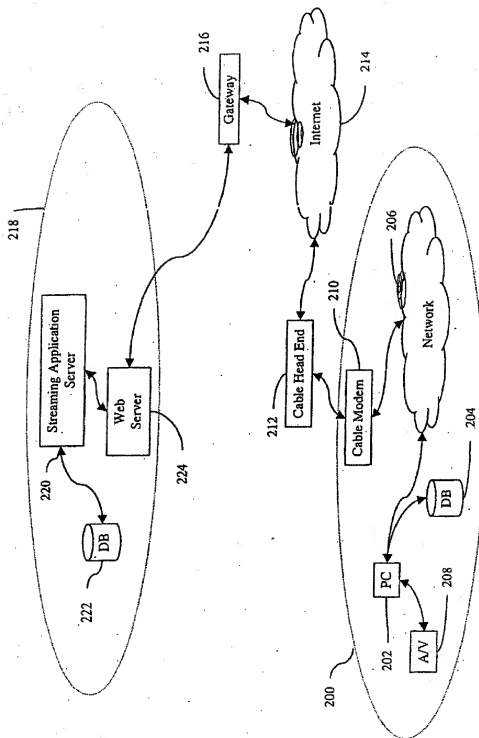


FIG. 2
Present Invention

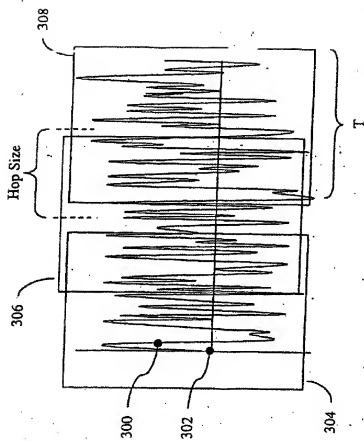


FIG. 3
Present Invention

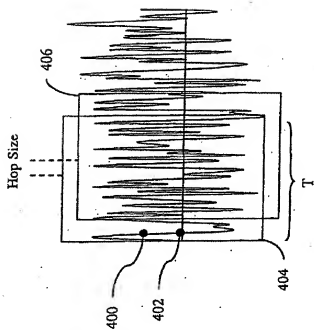


FIG. 4
Present Invention

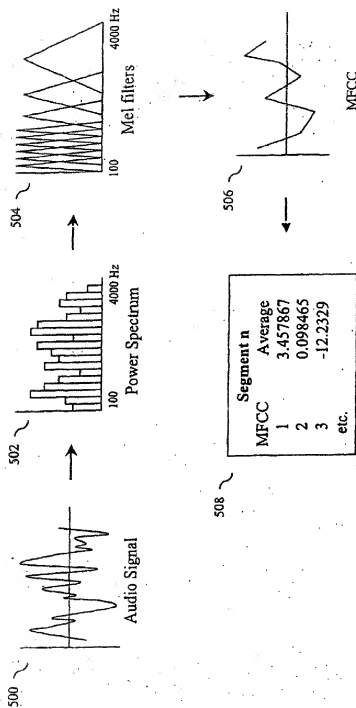


FIG. 5
Present Invention

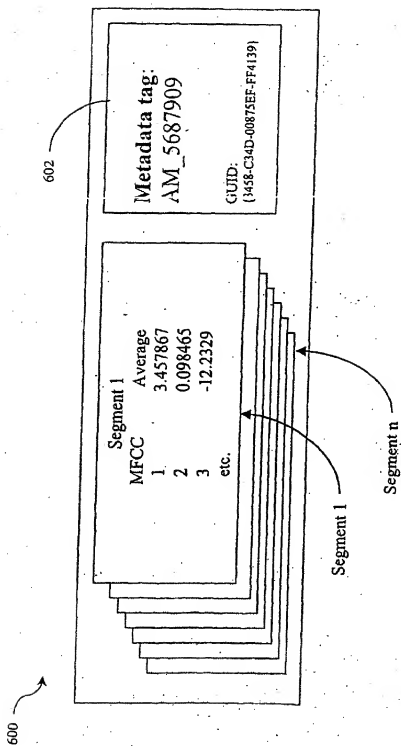


FIG. 6
Present Invention

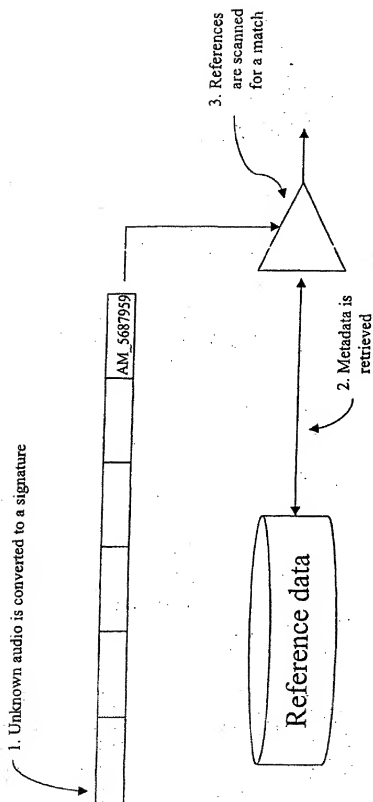


FIG. 7
Present Invention

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- US 5918223 A [0007]